

**Prioritätsbescheinigung über die Einreichung  
einer Patentanmeldung**

**Aktenzeichen:** ✓102 51 113.6

**Anmeldetag:** 2. November 2002

**Anmelder/Inhaber:** Philips Intellectual Property & Standards GmbH,  
Hamburg/DE  
(vormals: Philips Corporate Intellectual Property GmbH)

**Bezeichnung:** Verfahren zum Betrieb eines Spracherkennungssystems

**IPC:** G 10 L 15/00

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 28. Juli 2003  
Deutsches Patent- und Markenamt  
Der Präsident  
Im Auftrag

**PRIORITY DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

Stemp

## ZUSAMMENFASSUNG

### Verfahren zum Betrieb eines Spracherkennungssystems

Es wird ein Verfahren zum Betrieb eines Spracherkennungssystems beschrieben, in welchem ein Sprachsignal ( $S_I$ ) eines Benutzers detektiert und zur Erkennung einer im

5 Sprachsignal ( $S_I$ ) enthaltenen Sprachinformation analysiert wird. Hierbei ermittelt das Spracherkennungssystem einen Empfangsqualitätswert ( $S_Q$ ) oder einen Störwert, welcher eine aktuelle Empfangsqualität repräsentiert. Das Spracherkennungssystem schaltet in einen störunempfindlicheren Betriebsmodus um und/oder gibt ein

10 Warnsignal ( $S_W$ ) an den Benutzer aus, wenn der Empfangsqualitätswert ( $S_Q$ ) eine bestimmte Empfangsqualitätsschwelle unterschreitet bzw. der Störwert eine Störschwelle überschreitet. Darüber hinaus wird ein entsprechendes Spracherkennungssystem beschrieben.

BESCHREIBUNG

## Verfahren zum Betrieb eines Spracherkennungssystems

Die Erfindung betrifft ein Verfahren zum Betrieb eines Spracherkennungssystems, in welchem ein Sprachsignal eines Benutzers detektiert und zur Erkennung einer im Sprachsignal enthaltenen Sprachinformation analysiert wird. Darüber hinaus betrifft die Erfindung ein Spracherkennungssystem mit Mitteln zur Detektion eines Sprachsignals eines Benutzers und mit einer Spracherkennungseinrichtung, um das detektierte Sprachsignal zur Erkennung einer im Sprachsignal enthaltenen Sprachinformation zu analysieren.

In vielen Anwendungsbereichen werden zunehmend Spracherkennungssysteme genutzt, um eine sprachliche Kommunikation zwischen einem Benutzer und der jeweiligen Anwendung, z.B. einer Gerätesteuerung oder einem automatischen Informationsdienst, zu ermöglichen. Solche Sprach-Benutzerschnittstellen haben viele Vorteile. Zum einen erlauben sie ganz allgemein die freihändige Bedienung von Geräten. Zum anderen sind sie besonders als Benutzerschnittstellen für sehbehinderte oder körperbehinderte Personen geeignet, welche nicht in der Lage sind, optische Anzeigen zu lesen und/oder Eingabehilfsmittel wie eine Tastatur oder eine Maus zu bedienen und welche daher auf Sprache als Kommunikationsmittel angewiesen sind. Des Weiteren werden solche Benutzerschnittstellen zunehmend in automatischen Sprach-Dialogsystemen genutzt. Ein Beispiel für solche Sprach-Dialogsysteme sind automatische Anrufbeantwortungs- und Auskunftssysteme, wie sie inzwischen insbesondere von einigen größeren Firmen und Ämtern verwendet werden, um einen Anrufer möglichst schnell und komfortabel mit der gewünschten Information zu versorgen oder mit einer Stelle zu verbinden, welche für die speziellen Anfragen des Anrufers zuständig ist. Weitere Beispiele hierfür sind automatische Telefonauskunftssysteme, Fahrplanauskunftssysteme, Informationsdienste mit allgemeinen Veranstaltungshinweisen oder Kino- und Theaterprogrammen sowie beliebige Kombinationen der verschiedenen Auskunftssysteme.

Je nach Art der konkreten Anwendung, die das Spracherkennungssystem nutzt, erfolgt die technische Übermittlung des Sprachsignals des Benutzers zum Spracherkennungssystem über unterschiedliche Wege bzw. Geräte. Bei stationären Dialogsystemen, beispielsweise in Informationsterminals an öffentlichen Standorten wie Bahnhöfen, Flughäfen, Museen etc. weist das Spracherkennungssystem einen oder mehrere Empfangskanäle auf, an die geeignete Mikrofone, Headsets oder dgl. angeschlossen sind, über die die Nutzer mit dem Terminal bzw. dem Spracherkennungssystem kommunizieren können. Die zuvor beschriebenen automatischen Sprach-Dialogsysteme sind dagegen üblicherweise über ein Kommunikationsnetz, beispielsweise ein Telefonnetz oder das Internet, mittels entsprechender Endgeräte des Benutzers erreichbar. Hier wird folglich das Sprachsignal des Benutzers zunächst vom Endgerät erfasst, in elektrische Signale umgewandelt und über einen Sprach- und/oder Datenkanal des Kommunikationsnetzes an einen Anschluss des Sprachdialogsystems übermittelt.

- 15 Insbesondere bei solchen Spracherkennungssystemen, welche von den Benutzern mittels eines Kommunikationsendgeräts aus einer beliebigen Umgebung heraus kontaktiert werden können, ergibt sich zwangsläufig, dass die Empfangsqualität, mit der das Sprachsignal empfangen wird, stark differieren und sich sogar während einer Sitzung stark ändern kann. Die Empfangsqualität wird dabei zu einem wesentlichen Teil zum
- 20 einen durch die Qualität des Übertragungskanals, d. h. beispielsweise der Telefonverbindung, sowie zum anderen von dem Hintergrundgeräuschpegel bestimmt, der wiederum von der Umgebung abhängig ist, in der sich der Benutzer gerade befindet. Es ist klar, dass mit einer schlechteren Empfangsqualität auch die Arbeitsbedingungen für das Spracherkennungssystem ungünstiger sind, da Nebengeräusche oder auch kurzzeitige
- 25 Störungen bzw. Unterbrechungen des Übertragungskanals das Erkennungsergebnis verfälschen können. Mit zunehmender Verschlechterung der Empfangsqualität kommt es je nach Empfindlichkeit des Spracherkennungssystems früher oder später dazu, dass das Spracherkennungssystem kaum noch in der Lage ist, einigermaßen zufriedenstellende Erkennungsergebnisse zu liefern. Ungünstigerweise ist es dabei so, dass in der Regel

- gerade die Spracherkennungssysteme, welche möglichst schnell und natürlich auf eine Äußerung eines Benutzers reagieren und daher für den Benutzer an sich sehr komfortabel sind, besonders empfindlich auf Störungen reagieren. Dagegen sind Spracherkennungssysteme, welche z. B. nur zu bestimmten, genau vorgegebenen Zeitpunkten die
- 5 Eingabe von Sprachäußerungen des Benutzers zulassen, welche eine bestimmte, relativ hohe Lautstärke bei der Spracheingabe erfordern und welche relativ langsam auf Äußerungen reagieren, gegenüber Verschlechterungen der Empfangsqualität robuster.

Es ist Aufgabe der vorliegenden Erfindung, ein Verfahren zum Betrieb eines Spracherkennungssystems bzw. ein Spracherkennungssystem der eingangs genannten Art dahingehend zu verbessern, dass es auch bei einer Verschlechterung der Empfangsqualität mit möglichst guter Qualität weiter arbeitet und dabei den maximal möglichen Komfort für den Benutzer bietet.

- 15 Diese Aufgabe wird dadurch gelöst, dass ein Empfangsqualitätswert oder ein Störwert ermittelt wird, welcher eine aktuelle Empfangsqualität repräsentiert, und dass das Spracherkennungssystem in einen störungsempfindlicheren Betriebsmodus umschaltet und/oder ein Warnsignal an den Benutzer ausgibt, wenn der Empfangsqualitätswert eine bestimmte Empfangsqualitätsschwelle unterschreitet bzw. der Störwert eine Störschwelle überschreitet.
- 20

- Da die Beobachtung eines Empfangsqualitätswerts bezüglich des Unterschreitens einer bestimmten Empfangsqualitätsschwelle bis auf die Verwendung reziproker Werte sowie die entsprechende Umkehr der Grenzbedingung mit einer Beobachtung eines Störwerts,
- 25 beispielsweise die Höhe eines Hintergrundgeräuschsignals, bezüglich des Überschreitens einer bestimmten Störschwelle identisch ist, erfolgt die weitere Beschreibung der Erfindung – ohne die Erfindung in dieser Hinsicht zu beschränken – der Einfachheit halber in der Regel nur anhand der ersten Variante. Die einander entsprechenden Begriffe der beiden Varianten können in der nachfolgenden Beschreibung jederzeit gegen-
- 30 einander ausgetauscht werden.

Die Erfindung ermöglicht eine angemessene Reaktion des Spracherkennungssystems selbst und/oder des Benutzers auf eine Verschlechterung der Empfangsqualität, so dass folglich dem Benutzer immer der unter den aktuellen Empfangsbedingungen maximal mögliche Komfort geboten werden kann und andererseits sichergestellt ist, dass bei schlechter werdenden Empfangsbedingungen die Qualität der Erkennungsergebnisse nicht zu stark abfällt.

Als eine Reaktion des Spracherkennungssystems selbst ist dabei vorgesehen, dass das Spracherkennungssystem in einen robusteren Betriebsmodus umschaltet, in dem die Empfangsqualität sich weniger auf die Qualität der Spracherkennung auswirkt. Beispielsweise könnte ein Sprachaktivitätsdetektor, welcher das eingehende Sprachsignal zunächst detektiert, unempfindlicher geschaltet werden oder es können Wartezeiten, bis eine Reaktion des Spracherkennungssystems erfolgt, verlängert werden, so dass das Spracherkennungssystem insgesamt nicht mehr so schnell auf eventuelle Nebengeräusche reagiert. Alternativ oder zusätzlich wird ein Warnsignal, bei dem es sich bevorzugt um eine ausführlichere Warnmeldung in Form eines vom Spracherkennungssystem ausgegebenen Sprachsignals handelt, an den Benutzer ausgegeben, so dass dieser auch selbst auf die Verschlechterung der Empfangsqualität reagieren kann und sich beispielsweise in eine ruhigere Umgebung begibt oder bei Verwendung eines Mobilfunkgeräts versucht, die Mobilfunkverbindung durch Aufsuchen eines geeigneten Standorts zu verbessern.

Hierbei sind die unterschiedlichsten Kombinationen der beiden genannten Reaktionsmöglichkeiten denkbar. So kann beispielsweise nur ein Warnsignal an den Benutzer ausgegeben werden oder nur eine automatische Umschaltung in einen störungsunempfindlicheren Betriebsmodus durchgeführt werden. Es kann aber auch gleichzeitig mit der Umschaltung in den robusteren Betriebsmodus eine Warnmeldung an den Benutzer ausgegeben werden, die vorzugsweise den Hinweis enthält, dass der Betriebsmodus umgeschaltet wurde. Des Weiteren ist es auch möglich, dass zunächst nur das Warnsignal

an den Benutzer ausgegeben wird und beispielsweise nach Ablauf einer bestimmten Zeit, wenn die Empfangsqualität nicht wieder steigt oder sogar noch stärker absinkt, eine Umschaltung in den störungsunempfindlicheren Betriebsmodus erfolgt.

- 5 Die Empfangsqualitätsschwelle kann dabei auch in Abhängigkeit von bestimmten Parametern der jeweiligen Anwendung variabel festgelegt werden, beispielsweise in Abhängigkeit davon, welche Folgen eine fehlerhafte Erkennung eines Sprachsignals haben kann. Insbesondere ist es auch möglich, dass mehrere unterschiedliche Empfangsqualitätsschwellen festgelegt werden und in Abhängigkeit von diesen Empfangsqualitätsschwellen jeweils stufenweise unterschiedliche Betriebsmodi eingeschaltet und/oder bestimmte Warnsignale an den Benutzer ausgegeben werden.

- 15 Bei einem besonders bevorzugten Ausführungsbeispiel wird das Spracherkennungssystem automatisch in den vorherigen Betriebszustand zurückgeschaltet, wenn der Empfangsqualitätswert die Empfangsqualitätsschwelle wieder überschreitet. D. h. das Spracherkennungssystem wird beispielsweise automatisch wieder empfindlicher und reagiert somit für den Benutzer komfortabler, sobald sich der Benutzer aufgrund des Warnsignals erfolgreich um eine bessere Empfangsqualität bemüht, oder aus sonstigen Gründen die Empfangsqualität wieder besser wird.

- 20 Bei einem besonders bevorzugten Ausführungsbeispiel wird bei Unterschreiten der Empfangsqualitätsschwelle ein Barge-in-Betriebszustand des Spracherkennungssystems ausgeschaltet. D. h. das Spracherkennungssystem schaltet von einem Betriebsmodus, in dem der Benutzer die Möglichkeit hat, in Sprachausgaben des Spracherkennungssystems, sog. „Prompts“, hineinzureden, in einen „alternativen“ Betriebsmodus um, in  
25 welchem der Benutzer zunächst die vollständige Ausgabe des Prompts abwarten muss, bevor er auf das Prompt des Spracherkennungssystems reagieren kann. Solche Barge-in-Spracherkennungssysteme haben zum einen den Vorteil, dass sie gegenüber den „alternativen“ Systemen erheblich natürlicher, d. h. wie ein „richtiger“ Gesprächspartner, für

den Benutzer reagieren. Zum anderen kann ein erfahrener Benutzer ein solches Spracherkennungssystem erheblich schneller bedienen, da er oftmals die auszugebenden Prompts bereits kennt und daher die Ausgabe des Prompts durch eine vorzeitige Reaktion unterbrechen und somit den gesamten Dialog zeitlich verkürzen kann.

5

Die Ermittlung eines Empfangsqualitätswerts kann auf verschiedene Weise durchgeführt werden.

Besonders einfach ist ein Empfangsqualitätswert mittels eines Sprachaktivitätsdetektors ermittelbar, welcher ohnehin in der Regel am Eingang des Spracherkennungssystems dazu verwendet wird, um zunächst das eingehende Sprachsignal als solches zu detektieren und an die eigentliche Spracherkennungseinrichtung weiterzuleiten. Hierbei kann der Empfangsqualitätswert z. B. auf Basis eines in einer Sprechpause des Benutzers empfangenen Untergrundsignals bestimmt werden. D. h. es wird beispielsweise ein

- 15 "Rauschpegel bzw. die Basis-Signalenergie am Eingang innerhalb der Sprechpausen gemessen und als Maß für die Empfangsqualität genutzt. So kann beispielsweise festgestellt werden, ob der Benutzer sich in einer ruhigen oder lauten Umgebung befindet. Des Weiteren kann der Empfangsqualitätswert auch mit Hilfe der eigentlichen Spracherkennungseinrichtung selbst ermittelt werden, z. B. auf Basis von erreichten Konfi-
- 20 denzwerten für die Erkennungsergebnisse oder auf Basis von anderen Parametern, die beispielsweise von der Qualität des Erkennungsergebnisses oder vom Aufwand bei der Erkennung abhängen.

- 25 Die konkrete Verarbeitung des Empfangsqualitätswerts innerhalb des Spracherkennungssystems kann dann ebenfalls auf unterschiedliche Weise erfolgen.

- Zum einen kann der Sprachaktivitätsdetektor den Empfangsqualitätswert selbst beispielsweise an eine Dialog-Steuereinrichtung übermitteln. In diesem Fall kann z. B. ein Programmierer der jeweiligen Anwendung, welcher die Dialog-Steuereinrichtung ent-
- 30 sprechend seiner Anwendung ausbildet, festlegen, ab welcher Empfangsqualitäts-



schwelle das Spracherkennungssystem in welcher Weise reagiert bzw. welche Prompts an den Benutzer ausgegeben werden.

Alternativ oder zusätzlich ist es auch möglich, dass der Sprachaktivitätsdetektor ein  
5 Empfangsbeeinträchtigungssignal an die Dialog-Steuereinrichtung und/oder andere Komponenten des Systems ausgibt, wenn der Empfangsqualitätswert die Empfangsqualitätsschwelle unterschreitet. Die Dialog-Steuereinrichtung kann dann entsprechend auf das Empfangsbeeinträchtigungssignal reagieren und beispielsweise einen passenden Prompt an den Benutzer ausgeben. Außerdem kann über die Dialog-Steuereinrichtung und/oder die weiteren Systemkomponenten der Betriebszustand geändert werden.

Bei einem besonders komfortablen System wird ein eingehendes Signal detaillierter hinsichtlich der Art der Störung, die dafür verantwortlich ist, dass der Empfangsqualitäts-  
15 wert unterhalb der Empfangsqualitätsschwelle liegt, analysiert. Es wird dann als Warnsignal an den Benutzer ein Prompt ausgegeben, welches die entsprechenden Informationen enthält. D.h. es wird beispielsweise analysiert, ob es sich um eine schlechte Verbindung zwischen einem Kommunikationsendgerät des Benutzers und dem Spracherkennungssystem handelt oder ob die Hintergrundgeräusche in der Umgebung des Benutzers zu laut sind. Wenn der Benutzer die entsprechenden Informationen erhält, fällt es  
20 ihm leichter, in der richtigen Weise zu reagieren, um die Empfangsqualität zu verbessern. Diese Analyse kann u. a. mit Hilfe des Sprachaktivitätsdetektors und/oder der eigentlichen Spracherkennungseinrichtung erfolgen.

25 Ein erfindungsgemäßes Spracherkennungssystem muss zur Durchführung des erfindungsgemäßen Verfahrens zum einen eine entsprechende Qualitätskontrollereinrichtung zur Ermittlung eines entsprechenden Empfangsqualitätswerts oder eines Störwerts aufweisen. Zum anderen muss das Spracherkennungssystem einen Vergleich aufweisen, um den Empfangsqualitätswert mit einer bestimmten Empfangsqualitätsschwelle zu  
30 vergleichen oder um den Störwert mit einer bestimmten Störschwelle zu vergleichen.

Schließlich benötigt ein solches Spracherkennungssystem entsprechende Steuerungsmittel, beispielsweise eine Schalteinrichtung oder eine entsprechend programmierte Dialog-Steuereinrichtung, um das Spracherkennungssystem in einen störungsunempfindlicheren Betriebsmodus umzuschalten und/oder ein Warnsignal an den Benutzer auszugeben, wenn der Empfangsqualitätswert eine bestimmte Empfangsqualitätsschwelle unterschreitet bzw. der Störwert die Störschwelle überschreitet. Der Vergleich kann dabei auch in andere Komponenten des Systems, beispielsweise den Sprachaktivitätsdetektor oder die Dialog-Steuereinrichtung integriert sein.

Ein erfindungsgemäßes Spracherkennungssystem kann vorzugsweise in weiten Teilen mittels geeigneter Software auf einem Computer bzw. in einer Sprachsteuerung eines Geräts realisiert werden. So können z. B. die Spracherkennungseinrichtung und die Dialog-Steuereinrichtung vollständig in Form von Softwaremodulen realisiert werden.

Auch eine Einrichtung zur Generierung der geeigneten Prompts, beispielsweise ein sogenannter „TTS-Konverter“ (Text-To-Speech-Konverter) ist ebenfalls mittels geeigneter Software realisierbar. Alternativ ist es auch möglich, einen Prompt-Player zu verwenden, welcher bestimmte, im vorhinein aufgenommene Prompts für den Benutzer abspielt. Es ist natürlich erforderlich, dass das System entsprechend hardwaremäßig zu realisierende Möglichkeiten zur Spracheingabe und zur Sprachausgabe hat, beispielsweise ein Mikrofon und einen Lautsprecher oder eine Schnittstelle zum Anschluss an ein Kommunikationsnetz aufweist, über welches das Spracherkennungssystem mittels eines Endgeräts erreichbar ist.

Insbesondere ist es auch möglich, dass das Spracherkennungssystem nicht innerhalb eines einzelnen Geräts, beispielsweise auf einem einzelnen Server, realisiert ist, sondern dass verschiedene Komponenten des Systems an verschiedenen Orten angeordnet sind, welche über ein entsprechendes Netzwerk untereinander verbunden sind. Das erfindungsgemäße Spracherkennungssystem kann in Verbindung mit beliebigen Applikationen verwendet werden. Insbesondere kann es auch gleichzeitig, beispielsweise in Form eines Timesharing-Verfahrens, mehreren Applikationen als dienstleistendes

System zur Verfügung stehen und so für die verschiedenen Applikationen jeweils eine Benutzerschnittstelle bilden.

Die Erfindung wird im Folgenden unter Hinweis auf die beigelegten Figuren anhand  
5 zweier Ausführungsbeispiele näher erläutert. Es zeigen:

Figur 1 ein schematisches Blockdiagramm für ein erfindungsgemäßes Spracherkennungssystem gemäß einem ersten Ausführungsbeispiel,

Figur 2 ein schematisches Blockdiagramm für ein erfindungsgemäßes Spracherkennungssystem gemäß einem zweiten Ausführungsbeispiel.

Bei den in den beiden Figuren dargestellten Ausführungsbeispielen handelt es sich jeweils um sogenannte „barge-in-fähige“ Spracherkennungssysteme 1.  
15

An die Spracherkennungssysteme 1 ist jeweils zur Eingabe eines Sprachsignals  $S_1$  ein Mikrofon 2 angeschlossen. Zur Ausgabe von akustischen Signalen bzw. Sprachausgaben (Prompts) an den Benutzer ist an die Spracherkennungssysteme 1 jeweils ein Lautsprecher 3 angeschlossen. Alternativ können sich das Mikrofon 2 und der Lautsprecher 3 auch in einem vom Spracherkennungssystem 1 selbst entfernten Endgerät befinden, welches über eine geeignete Schnittstelle, beispielsweise ein Telefonnetz o. Ä., mit dem Spracherkennungssystem 1 verbunden ist.  
20

Ein vom Mikrofon 2 aus eingehendes Signal gelangt bei beiden Spracherkennungssystemen 1 jeweils zunächst zu einem Echofilter 4 (auch „Echo Cancellator“ genannt). Dieser Echofilter 4 dient dazu, um in den Fällen, in denen der Benutzer gleichzeitig spricht, obwohl noch ein Prompt  $S_0$  durch das Spracherkennungssystem 1 selbst ausgegeben wird, das Echo des Prompts  $S_0$ , welches in gedämpfter Form dem Sprachsignal  $S_1$  überlagert sein kann, aus dem Eingangssignal auszufiltern.  
25  
30

Das Sprachsignal  $S_1$  wird dann an einen Sprachaktivitätsdetektor 5 weitergeleitet, welcher die Signalenergie und somit die Sprachaktivität des Benutzers misst. Dieser Sprachaktivitätsdetektor 5 dient zum einen dazu, um den Beginn und das Ende eines Sprachsignals sowie Sprechpausen innerhalb einer Äußerung des Benutzers zu detek-  
5 tieren. Zum anderen gibt dieser Sprachaktivitätsdetektor ein Barge-in-Signal  $S_{BI}$  aus, welches zu einem Text-to-Speech-Konverter 8 (im Folgenden TTS-Konverter 8 genannt) und an eine Dialog-Steuereinrichtung 10 geleitet wird, um in den Fällen, in denen ein Benutzer in einen ausgegebenen Prompt hinein spricht, die weitere Ausgabe des Prompts durch den TTS-Konverter 8 zu verhindern.

Im vorliegenden Fall befindet sich in der Ausgangsleitung für das Barge-in-Signal  $S_{BI}$  des Sprachaktivitätsdetektors 5 eine Barge-in-Schalteinheit 9. Nur im geschlossenen Zustand dieser Barge-in-Schalteinheit 9 ist das System 1 tatsächlich barge-in-fähig. Bei geöffnetem Schalter (wie in Figur 1 dargestellt) reagiert das Spracherkennungssystem 1  
15 dagegen auf eine Spracheingabe des Benutzers während der Ausgabe eines Prompts nicht.

Vom Sprachaktivitätsdetektor 5 aus wird das Sprachsignal  $S_1$  selbst an eine automatische Spracherkennungseinrichtung 7 weitergeleitet, in welcher die eigentliche Analyse  
20 des Sprachsignals  $S_1$  zur Erkennung der darin enthaltenen Sprachinformation durchgeführt wird. Das Erkennungsergebnis wird dann in Textform bzw. in einer anderen maschinenlesbaren Form an die Dialog-Steuereinrichtung 10 weitergeleitet.

Diese Dialog-Steuereinrichtung 10 reagiert schließlich in vorbestimmter Weise, d. h.  
25 gemäß einem vorgegebenen Dialogablauf, auf das Erkennungsergebnis und sucht beispielsweise aus einem Speicher 12 eine vom Benutzer gewünschte Information heraus oder hinterlegt dort vom Benutzer eingegebene Daten. Außerdem steuert diese Dialog-Steuereinrichtung den TTS-Konverter 8 und übergibt diesem die Daten für die an den

Benutzer auszugebenden Prompts. Die Prompts werden dann vom TTS-Konverter 8 aus über den Echofilter 4 an den Lautsprecher 3 übermittelt und dort ausgegeben. Der genaue Aufbau der Dialog-Steuereinrichtung 10 ist in der Regel abhängig von der jeweiligen Applikation.

5

Erfindungsgemäß weisen die beiden in den Figuren dargestellten Ausführungsbeispiele jeweils neben einer Barge-in-Schalteinheit 9 eine Qualitätskontrolleinrichtung 6 auf, welche Teil des Sprachaktivitätsdetektors 5 ist. In dieser Qualitätskontrolleinrichtung 6 wird ein Empfangsqualitätswert  $S_Q$  ermittelt. Hierbei kann es sich beispielsweise um den reziproken Wert eines Hintergrundgeräuschpegels handeln, welcher vor Beginn der Äußerung und/oder in den Sprechpausen des Benutzers vom Sprachaktivitätsdetektor 5 gemessen wird.

15

Der Empfangsqualitätswert  $S_Q$  wird an die Dialog-Steuereinrichtung 10 übermittelt, welche den Empfangsqualitätswert  $S_Q$  innerhalb eines Vergleichers 11, der z. B. in Form eines Software-Moduls realisiert ist, mit einer vorgegebenen Empfangsqualitätschwelle vergleicht. Sobald der Empfangsqualitätswert  $S_Q$  unterhalb der vorgegebenen Schwelle liegt, wird von der Dialog-Steuereinrichtung 10 ein Barge-in-Ausschaltsignal  $S_{BA}$  an die Barge-in-Schalteinheit 9 übermittelt und dort der Schalter in den in Figur 1 dargestellten Zustand gebracht. D. h. das Spracherkennungssystem 1 wird von einem barge-in-fähigen Betriebszustand in einen nicht barge-in-fähigen Betriebszustand umgeschaltet. Dies hat den Vorteil, dass es bei zu hoher Hintergrundgeräuschschwelle nicht dazu kommen kann, dass das Spracherkennungssystem 1 während einer Ausgabe eines Prompts ständig aufgrund von Hintergrundgeräuschen, welche nicht dem Benutzer zugeordnet sind und nicht zur Unterbrechung des Prompts dienen sollen, unterbrochen wird. D. h. das Spracherkennungssystem 1 schaltet in einen Betriebszustand um, der zwar zu einer etwas unnatürlicheren und damit für den Benutzer unbequemerem Dialogweise führt, der aber erheblich robuster und für die zu diesem Zeitpunkt vorliegenden Empfangsbedingungen erheblich geeigneter ist.

20

25

30

Zusätzlich lässt die Dialog-Steuereinrichtung 10, sobald der Empfangsqualitätswert  $S_Q$  unterhalb der vorgegebenen Empfangsqualitätsschwelle liegt, ein entsprechendes Warnsignal in Form eines geeigneten Warn-Prompts  $S_W$  vom TTS-Generator 8 erzeugen und ausgeben, beispielsweise eine Ansage „Die Hintergrundgeräusche sind zu laut. Bitte  
5 warten Sie die Ausgabe eines Prompts vollständig ab, bevor Sie antworten.“

Die Dialog-Steuereinrichtung 10 ist derart eingerichtet, dass sie automatisch, wenn der Empfangsqualitätswert  $S_Q$  wieder oberhalb der vorgegebenen Empfangsqualitätsschwelle liegt, das Barge-in-Ausschaltsignal  $S_{BA}$  aufhebt und somit der Schalter innerhalb der Barge-in-Schalteinheit 9 wieder geschlossen wird.

Grundsätzlich steht es dabei im Ermessen des Programmierers der Dialog-Steuereinrichtung 10, beliebige Empfangsqualitätsschwellen festzulegen, ab denen ein Barge-in-Zustand ausgeschaltet wird und/oder bei denen eine Warnung an den Benutzer erfolgt,  
15 damit dieser sich gegebenenfalls bemüht, eine ruhigere Umgebung aufzusuchen, sofern dies möglich ist.

Figur 2 zeigt ein geringfügig anderes Ausführungsbeispiel eines erfindungsgemäßen Spracherkennungssystems 1.

20 Bei diesem Ausführungsbeispiel erfolgt der Vergleich des Empfangsqualitätswerts  $S_Q$  mit der Empfangsqualitätsschwelle direkt in der Qualitätskontrolleinrichtung 6, d. h. der Vergleich 11 ist in die Qualitätskontrolleinrichtung 6 integriert. Die Qualitätskontrolleinrichtung 6 gibt bei Unterschreiten der vorgegebenen Empfangsqualitätsschwelle ein  
25 Empfangsbeeinträchtigungsanzeigesignal  $S_{EB}$  aus, welches gleichzeitig die Barge-in-Schalteinheit 9 ausschaltet und an die Dialog-Steuereinrichtung 10 gegeben wird, damit dort je nach Programmierung die Ausgabe eines geeigneten Warn-Prompts  $S_W$  über den TTS-Konverter 8 veranlasst wird.

16

Es wird noch einmal darauf hingewiesen, dass es sich bei den in den Figuren dargestellten Spracherkennungssystemen 1 nur um spezielle Ausführungsbeispiele der Erfindung handelt und es im Ermessen des Fachmanns steht, die Spracherkennungssysteme 1 bzw. das Verfahren zum Betrieb der Spracherkennungssysteme 1 im Rahmen der Erfindung auf verschiedene Weise zu modifizieren.

So ist es insbesondere möglich, dass beispielsweise die Qualitätskontrolleinheit 6 auch Teil der Dialog-Steuereinrichtung 10 ist. Ebenso kann die Barge-in-Schalteneinrichtung 9 auch Teil des Sprachaktivitätsdetektors 5 oder der Dialog-Steuereinrichtung 10 sein und beispielsweise auch als rein softwaremäßiger Schalter ausgebildet sein. Des Weiteren ist es möglich, dass die Qualitätskontrolleinheit 6 rein hardwaremäßig als eigenständiges Modul aufgebaut ist. Außerdem kann das Spracherkennungssystem zusätzliche Komponenten bzw. weitere Funktionen aufweisen, beispielsweise ein intelligentes, grammatikbasiertes Barge-in-Verfahren durchführen. Hierbei wird nicht bereits von dem Sprachaktivitätsdetektor das Barge-In-Signal erzeugt, sondern erst von der Spracherkennungseinrichtung, welche zunächst prüft, ob ein eingehendes Signal wirklich einen sinnvollen Inhalt hat und daher als Barge-In-Fall anzusehen ist. Weitere Variationen können beispielsweise eine Sprechererkennung oder andere Erweiterungen des Spracherkennungssystems enthalten.

## PATENTANSPRÜCHE

1. Verfahren zum Betrieb eines Spracherkennungssystems (1), in welchem ein Sprachsignal ( $S_1$ ) eines Benutzers detektiert und zur Erkennung einer im Sprachsignal ( $S_1$ ) enthaltenen Sprachinformation analysiert wird;

dadurch gekennzeichnet,

dass ein Empfangsqualitätswert ( $S_Q$ ) oder ein Störwert ermittelt wird, welcher eine aktuelle Empfangsqualität repräsentiert, und dass das Spracherkennungssystem (1) in einen störunempfindlicheren Betriebsmodus umschaltet und/oder ein Warnsignal ( $S_w$ ) an den Benutzer ausgibt, wenn der Empfangsqualitätswert ( $S_Q$ ) eine bestimmte Empfangsqualitätsschwelle unterschreitet bzw. der Störwert eine Störschwelle überschreitet.

2. Verfahren nach Anspruch 1,

dadurch gekennzeichnet,

dass das Spracherkennungssystem automatisch in den vorherigen Betriebszustand zurückschaltet, wenn der Empfangsqualitätswert ( $S_Q$ ) die Empfangsqualitätsschwelle wieder überschreitet bzw. der Störwert die Störschwelle wieder unterschreitet.

3. Verfahren nach Anspruch 1 oder 2,

dadurch gekennzeichnet,

dass bei Unterschreiten der Empfangsqualitätsschwelle bzw. Überschreiten der Störschwelle ein Barge-in-Betriebszustand des Spracherkennungssystems (1) ausgeschaltet wird.



18

4. Verfahren nach einem der Ansprüche 1 bis 3,

dadurch gekennzeichnet,

dass der Empfangsqualitätswert ( $S_Q$ ) oder Störwert mittels eines Sprachaktivitätsdetektors (5) ermittelt wird.

5

5. Verfahren nach einem der Ansprüche 1 bis 4,

dadurch gekennzeichnet,

dass der Empfangsqualitätswert ( $S_Q$ ) oder Störwert auf Basis eines vor Beginn der Äußerung und/oder in einer Sprechpause des Benutzers empfangenen Untergrundsignals ermittelt wird.

6. Verfahren nach Anspruch 4 oder 5,

dadurch gekennzeichnet,

15 dass der Sprachaktivitätsdetektor (5) den Empfangsqualitätswert ( $S_Q$ ) oder Störwert selbst und/oder bei Unterschreiten der Empfangsqualitätsschwelle bzw. bei Überschreiten der Störschwelle ein Empfangsbeeinträchtigungssignal ( $S_{EB}$ ) an eine Dialog-Steuereinrichtung (10) übermittelt.

7. Verfahren nach einem der Ansprüche 1 bis 6,

dadurch gekennzeichnet,

20 dass die Dialog-Steuereinrichtung (10) bei Empfang des Empfangsbeeinträchtigungssignals ( $S_{EB}$ ) und/oder wenn der empfangene Empfangsqualitätswert ( $S_Q$ ) die Empfangsqualitätsschwelle unterschreitet bzw. der Störwert die Störschwelle überschreitet, die Ausgabe eines Prompts ( $S_w$ ) an den Benutzer auslöst, welcher die  
25 Information enthält, dass die Empfangsbedingungen schlecht sind.

8. Verfahren nach einem der Ansprüche 1 bis 7,

dadurch gekennzeichnet,

dass ein eingehendes Signal hinsichtlich der Art der Störung, die dafür verantwortlich ist, dass der Empfangsqualitätswert ( $S_Q$ ) unterhalb der Empfangsqualitätsschwelle bzw. der Störwert oberhalb der Störschwelle liegt, analysiert wird, und an den Benutzer ein Prompt ( $S_w$ ) ausgegeben wird, welches diese Information enthält.

9. Spracherkennungssystem (1) mit Mitteln (5) zur Detektion eines Sprachsignals ( $S_i$ ) eines Benutzers und einer Spracherkennungseinrichtung (7), um das detektierte Sprachsignal ( $S_i$ ) zur Erkennung einer im Sprachsignal enthaltenen Sprachinformation zu analysieren,

gekennzeichnet durch,

eine Qualitätskontrollereinrichtung (6) zur Ermittlung eines Empfangsqualitätswerts ( $S_Q$ ) oder eines Störwerts, welcher eine aktuelle Empfangsqualität repräsentiert,

einen Vergleich, um den Empfangsqualitätswert ( $S_Q$ ) mit einer bestimmten Empfangsqualitätsschwelle zu vergleichen oder um den Störwert mit einer bestimmten Störschwelle zu vergleichen,

und Steuerungsmittel (9, 10), welche derart ausgebildet sind, dass das

Spracherkennungssystem (1) in einen störungsempfindlicheren Betriebsmodus

umgeschaltet wird und/oder ein Warnsignal ( $S_w$ ) an den Benutzer ausgegeben wird,

wenn der Empfangsqualitätswert die Empfangsqualitätsschwelle unterschreitet bzw. der Störwert die Störschwelle überschreitet.

10. Computerprogramm mit Programmcode-Mitteln, um alle Schritte eines Verfahrens

nach einem der Ansprüche 1 bis 9 auszuführen, wenn das Programm auf einem Computer ausgeführt wird.

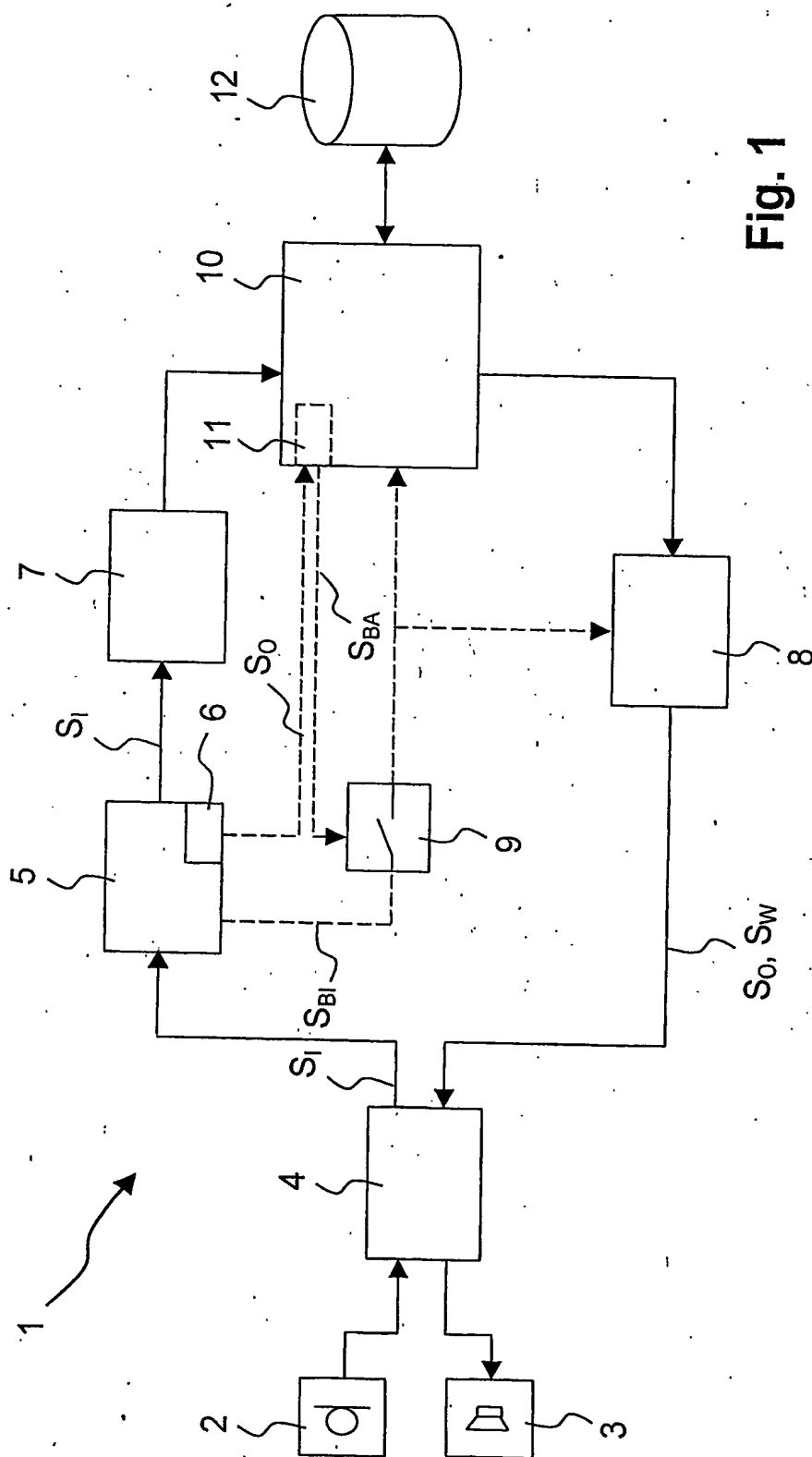


Fig. 1

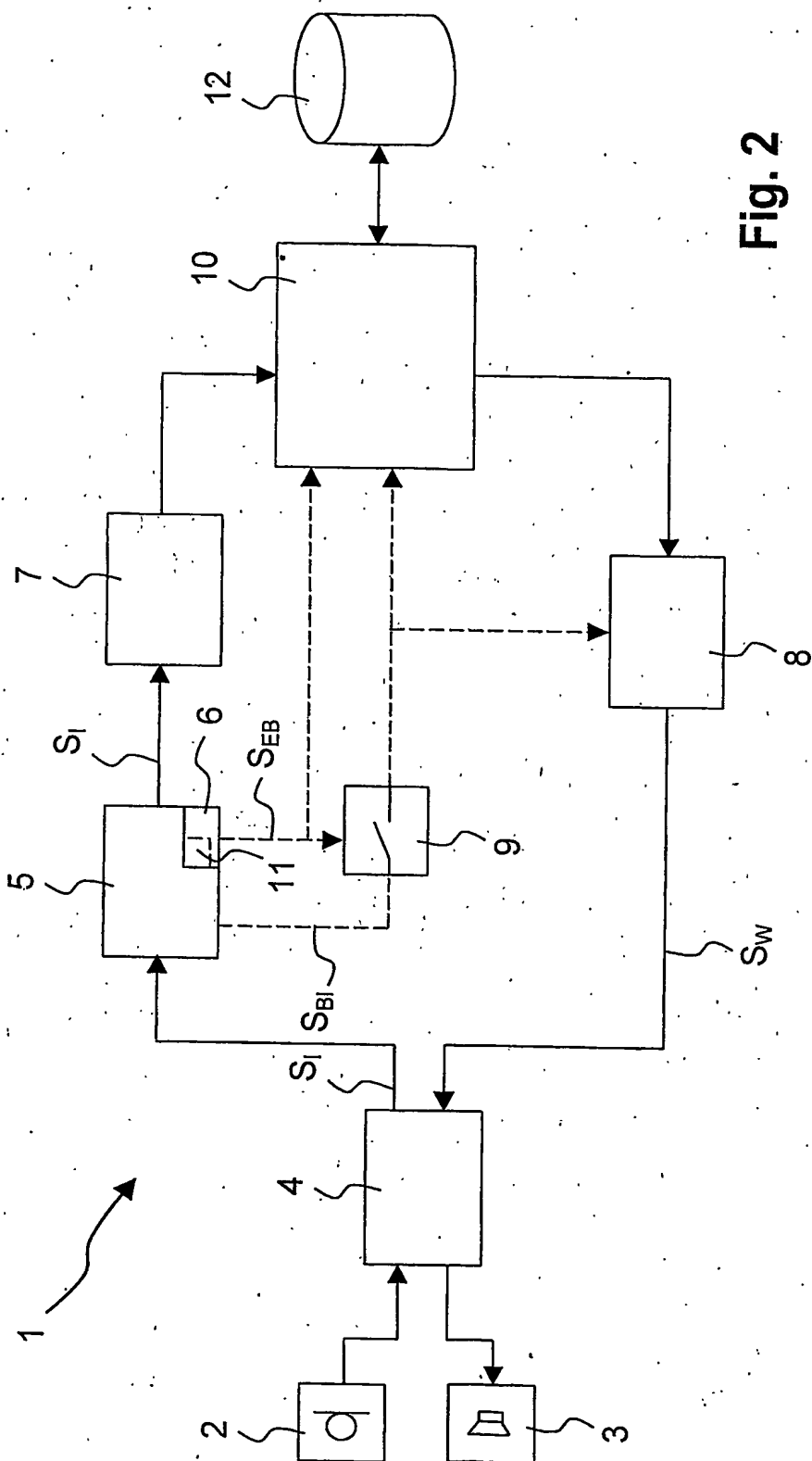


Fig. 2

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record.**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**